WHAT IS CLAIMED IS:

1. A mathematical expression recognizing device comprising:

a character recognition unit configured to recognize characters in a document image containing a text and a mathematical expression;

a first dictionary configured to store a pair of evaluation scores for each type of word that can be identified by means of normal expression, the score showing the possibility of belonging to the text and that of belonging to the mathematical expression;

an evaluation unit configured to obtain the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words included in the characters recognized by the character recognition unit with reference to the first dictionary; and

a mathematical expression detecting unit configured to search for an optimal path connecting words by selecting one of the text and the mathematical expression based on a formative grammar and the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words, thereby detecting characters belonging to the mathematical expression.

2. The device according to claim 1, wherein said mathematical expression detection unit comprises:

a second dictionary configured to store a connectable a part of speech and mathematical expression as the formative grammar; and

a search unit configured to search for a path connecting the words and showing the largest evaluation score given to the word as the mathematical expression or the text out of all possible inter-word connection paths as the optimal path, by selecting either the text or mathematical expression for each word according to the part of speech of the word and the formative grammar read out from said second dictionary.

3. The device according to claim 1, further comprising:

a memory configured to store a plurality of items of sample information indicating a relation of a normalization size and a center position between each pair of consecutively arranged characters in terms of the types of the characters including a horizontal positional relationship, character/subscript relationship and character/superscript relationship; and

a determination unit configured to calculate the relation of the normalization size and the center position between each pair of consecutively arranged characters included in the mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship

based on the calculated relation of the normalization size and the center position and the sample information corresponding to the calculated relation of the types of the two consecutively arranged characters.

4. The device according to claim 3, further comprising:

a memory configured to storing a global evaluation condition for determined based on the distribution of the heights of the characters contained in said mathematical expression region; and

a unit for configured to search for an optimal path for connecting the characters in each of said mathematical expression regions without contradiction, select an inter-character structure candidate having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship for each pair of consecutively arranged characters based on said global evaluation condition and said link candidates, and recognize the horizontal positional relationship, the character/subscript relationship or the character/superscript relationship of said pair of consecutively arranged characters based on the result of the search operation.

5. The device according to claim 4, wherein said global evaluation condition comprises at least one of the relationship of the height of a character contained in a subscript region and the height of each of other

characters, the positional relationship between a base line and a character contained in the subscript region and the dispersion of heights among characters located on the same horizontal level.

6. The device according to claim 3, further comprising:

a decomposing unit configured to decompose each mathematical expression detected by said mathematical expression detection unit into components and remove at least left indexes, accent marks, root signs, and dots from each component, and wherein said determination unit obtains link candidates for the components from which the left indexes, accent marks, root signs, or dots is removed.

7. A mathematical expression recognizing device comprising:

a character recognition unit configured to recognize characters in a document image containing a text and a mathematical expression;

a detecting unit configured to detect a mathematical expression region from the characters recognized by the character recognition unit:

a memory configured to store a plurality of items of sample information indicating a relation of a normalization size and a center position between each pair of consecutively arranged characters in terms of the types of the characters including a horizontal

positional relationship, character/subscript relationship and character/superscript relationship; and

a unit configured to calculate the relation of the normalization size and the center position between each pair of consecutively arranged characters included in the mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship based on the calculated relation of the normalization size and the center position and the sample information corresponding to the calculated relation of the types of the two consecutively arranged characters.

8. A mathematical expression recognizing device comprising:

a character recognition unit configured to recognize characters in a document image containing a text and a mathematical expression;

a detecting unit configured to detect a mathematical expression region from the characters recognized by the character recognition unit;

a memory configured to store a plurality of items of sample information indicating a relation of a normalization size and a center position between each pair of consecutively arranged characters in terms of the types of the characters including a horizontal positional relationship, character/subscript

relationship and character/superscript relationship;

a unit configured to calculate the relation of the normalization size and the center position between each pair of consecutively arranged characters included in the mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship based on the calculated relation of the normalization size and the center position and the sample information corresponding to the calculated relation of the types of the two consecutively arranged characters; and

a unit configured to search for an optimal path for connecting the characters in each of said mathematical expression regions without contradiction, select an inter-character structure candidate having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship for each pair of consecutively arranged characters based on said global evaluation condition and said link candidates, and recognize the horizontal positional relationship, the character/subscript relationship or the character/superscript relationship of said pair of consecutively arranged characters based on the result of the search operation.

9. A mathematical expression recognizing method comprising:

recognizing characters in a document image containing a text and a mathematical expression;

referring to a first dictionary which stores a pair of evaluation scores for each type of word that can be identified by means of normal expression, the score showing the possibility of belonging to the text and that of belonging to the mathematical expression to obtain the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words included in the characters recognized by the character; and

searching for an optimal path connecting words by selecting one of the text and the mathematical expression based on a formative grammar and the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words, thereby detecting characters belonging to the mathematical expression.

10. A mathematical expression recognizing method comprising:

recognizing characters in a document image containing a text and a mathematical expression;

detecting a mathematical expression region from the recognized characters;

referring to a memory which stores a plurality of items of sample information indicating a relation of a normalization size and a center position between each

pair of consecutively arranged characters in terms of the types of the characters including a horizontal positional relationship, character/subscript relationship and character/superscript relationship; and

calculating the relation of the normalization size and the center position between each pair of consecutively arranged characters included in the mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship based on the calculated relation of the normalization size and the center position and the sample information corresponding to the calculated relation of the types of the two consecutively arranged characters.

11. A mathematical expression recognizing method comprising:

recognizing characters in a document image containing a text and a mathematical expression;

detecting a mathematical expression region from the recognized characters;

referring to a memory which stores a plurality of items of sample information indicating a relation of a normalization size and a center position between each pair of consecutively arranged characters in terms of the types of the characters including a horizontal

positional relationship, character/subscript relationship and character/superscript relationship;

calculating the relation of the normalization size and the center position between each pair of consecutively arranged characters included in the mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship based on the calculated relation of the normalization size and the center position and the sample information corresponding to the calculated relation of the types of the two consecutively arranged characters; and

searching for an optimal path for connecting the characters in each of said mathematical expression regions without contradiction, selecting an inter-character structure candidate having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship for each pair of consecutively arranged characters based on said global evaluation condition and said link candidates, and recognizing the horizontal positional relationship, the character/subscript relationship or the character/superscript relationship of said pair of consecutively arranged characters based on the result of the search operation.

12. A character recognizing device for reading

a document including a mathematical expression and recognizing respectively a text region and a mathematical expression region, comprising:

a character recognition unit configured to recognize the document image including the mathematical expression;

a first dictionary configured to store a pair of evaluation scores for each type of word that can be identified by means of normal expression, the score showing the possibility of belonging to the text and that of belonging to the mathematical expression; and

an evaluation unit configured to obtain the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words included in the characters recognized by the character recognition unit with reference to the first dictionary, search for an optimal path connecting words by selecting one of the text and the mathematical expression based on a formative grammar and the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words, thereby discriminating between the text and the mathematical expression.

13. A character recognizing device for reading a document including a mathematical expression and recognizing respectively a text region and

a mathematical expression region, comprising:

a character recognition unit configured to recognize the document image including the mathematical expression;

a detecting unit configured to detect a mathematical expression region from the characters recognized by the character recognition unit;

a memory configured to store a plurality of items of sample information indicating a relation of a normalization size and a center position between each pair of consecutively arranged characters in terms of the types of the characters including a horizontal positional relationship, character/subscript relationship and character/superscript relationship; and

a unit configured to calculate the relation of the normalization size and the center position between each pair of consecutively arranged characters included in the mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship based on the calculated relation of the normalization size and the center position and the sample information corresponding to the calculated relation of the types of the two consecutively arranged characters, thereby recognizing a mathematical expression.

14. A mathematical expression recognizing device

comprising:

a character recognizing unit configured to recognize characters in a document image containing a mathematical expression;

a unit configured to detect a mathematical expression region from the outcome of character recognition obtained by said character recognizing means;

a unit configured to store a plurality of pieces of sample information on the inter-character relationship of the sizes of normalization and that of the center positions of each pair of consecutively arranged characters in terms of the types of the characters and positional relationships of horizontal positional relationship, an inter-character relationship determining unit configured to computationally determining the relationship of the sizes of normalization and that of the center positions of each of all the pairs of consecutively arranged characters in a mathematical expression region and obtain link candidates as combinations of inter-character structure candidates showing the respective possibilities of having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship based on the result of computation and the sample information and their respective evaluation scores;

a unit configured to store a global evaluation condition based on the distribution of the heights of the characters contained in said mathematical expression regions; and

a unit configured to search for an optimal path for connecting the characters in each of said mathematical expression regions without contradiction, select an inter-character structure candidate having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship for each pair of consecutively arranged characters, and recognize the horizontal positional relationship, the character/subscript relationship or the character/superscript relationship of said pair of consecutively arranged characters based on the result of the search operation.

15. A character recognizing method for reading a document including a mathematical expression and recognizing respectively a text region and a mathematical expression region, comprising:

recognizing the document image including the mathematical expression;

referring to a first dictionary which stores a pair of evaluation scores for each type of word that can be identified by means of normal expression, the score showing the possibility of belonging to the text and that of belonging to the mathematical expression;

and

obtaining the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words included in the characters recognized by the character recognition unit with reference to the first dictionary, searching for an optimal path connecting words by selecting one of the text and the mathematical expression based on a formative grammar and the evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words, thereby discriminating between the text and the mathematical expression.

16. A character recognizing method for reading a document including a mathematical expression and recognizing respectively a text region and a mathematical expression region, comprising:

recognizing the document image including the mathematical expression;

detecting a mathematical expression region from the characters recognized by the character recognition unit;

referring a memory which stores a plurality of items of sample information indicating a relation of a normalization size and a center position between each pair of consecutively arranged characters in terms of the types of the characters including a horizontal

positional relationship, character/subscript relationship and character/superscript relationship; and

calculating the relation of the normalization size and the center position between each pair of consecutively arranged characters included in the mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship based on the calculated relation of the normalization size and the center position and the sample information corresponding to the calculated relation of the types of the two consecutively arranged characters, thereby recognizing a mathematical expression.

17. A mathematical expression recognizing method comprising:

recognizing characters in a document image containing a mathematical expression;

detecting a mathematical expression region from the outcome of character recognition obtained by said character recognizing means;

referring to a memory which stores pieces of sample information on the inter-character relationship of the sizes of normalization and that of the center positions of each pair of consecutively arranged characters in terms of the types of the characters and positional relationships of horizontal positional

relationship, an inter-character relationship determining unit configured to computationally determining the relationship of the sizes of normalization and that of the center positions of each of all the pairs of consecutively arranged characters in a mathematical expression region and obtaining link candidates as combinations of inter-character structure candidates showing the respective possibilities of having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship based on the result of computation and the sample information and their respective evaluation scores;

referring to a memory which stores a global evaluation condition based on the distribution of the heights of the characters contained in said mathematical expression regions; and

searching for an optimal path for connecting the characters in each of said mathematical expression regions without contradiction, selecting an inter-character structure candidate having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship for each pair of consecutively arranged characters, and recognizing the horizontal positional relationship, the character/subscript relationship or the character/superscript relationship of said pair of consecutively

arranged characters based on the outcome of the search operation.